**Trabalho Final BIE5782**

**Proposta A**

Função plot.coverage

Função irá criar um gráfico detalhando a cobertura do sequenciamento por base de uma ou várias amostras sequenciadas por exoma. O eixo X terá os números de pares de bases, enquanto o eixo Y tem a cobertura base a base, e uma linha pelo gráfico indicando os pontos da cobertura.

A importância de se ter uma cobertura boa de cada base é a confiança do que se está vendo é real ou não, ou seja, que a presença de variante, indels ou até mesmo a falta de variantes na sequencia esteja realmente no material genético do paciente. Muitos laboratórios padronizam o mínimo de cobertura desejável como 20 reads por base, e pedaços do genoma que não alcançam essa cobertura acabam por ter de ser refeitos, o que gera um grande custo para o laboratório. A ideia desse gráfico, então, é de facilitar a ilustração e, consequentemente, a visualização da quantidade de reads por base em um gene, ou seja, a cobertura do sequenciamento. Será tomado como exemplo os gráficos de cobertura do site do ExAC (http://exac.broadinstitute.org/).

Todas as informações necessárias para a criação do gráfico serão retiradas de um arquivo .BED provindo de um .BAM que contenha as informações de cobertura. Este arquivo possui formato de tabela, universal para o tipo de arquivo, e poderá ser lido como um data frame. Cada uma das colunas possui uma informação importante para a estimativa dessa cobertura por base, sendo elas o cromossomo, início do transcrito, final do transcrito, nome do gene, tamanho do fragmento, base do fragmento e cobertura na base.

O gene a ser plotado também poderá ser escolhido pelo usuário como argumento da função. A função também terá como opção fazer o gráfico de apenas uma amostra, que no caso só iria plotar as informações presentes no arquivo, dependendo do gene escolhido pela pessoa (que deve estar no arquivo de origem, se não a função retornará erro), ou então fazer o gráfico de cobertura de diversas amostras (que entrarão em um for), podendo serem feitas a média (mean()), mediana (median()) ou variancia (var()) por base, dependendo da escolha do usuário entre cada uma ou todas as três, que serão então plotadas em gráficos diferentes.

**Proposta B**

Função plot.saturation

Função irá criar um gráfico de saturação de cobertura por quantidade total de reads na amostra, para responder a questão de quantos reads são necessários para se atingir uma média/mediana de 100 vezes de cobertura. O gráfico então terá o eixo X com a quantidade de reads em milhões, e o eixo Y com a média/mediana de 0 a 100. Para isso será necessário como entrada uma tabela com a informação de cobertura com a quantidade de reads desejados para várias amostras, a fim de criar o gráfico. Cada uma das amostras vira um ponto nos gráficos, que serão unidos por uma linha. A ideia do gráfico é ajudar o pesquisador a decidir a quantidade de reads necessários para o sequenciamento chegar a uma cobertura ideal, para que informações importantes como por exemplo variantes pontuais, não sejam perdidas na análise.